

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
(A) NAME: Kossmann, Jens
Buttcher, Volker
Welsh, Thomas
- (ii) TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE
OF FACILITATING THE SYNTHESIS OF LINEAR
ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND
MICROORGANISMS
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
(B) STREET: 1251 Avenue of the Americas
(C) CITY: New York
(D) STATE: New York
(E) COUNTRY: United States of America
(F) ZIP: 10020
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: DE P 44 17 879.4
(B) FILING DATE: 18-MAY-1994
- (vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: DE P 44 47 388.5
(B) FILING DATE: 22-DEC-1994
- (vii) ATTORNEY/AGENT INFORMATION:
(A) NAME: James F. Haley, Jr., Esq.
(B) REGISTRATION NUMBER: 27,794
(C) DOCKET NUMBER: GFB-1
- (viii) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212) 596-9000
(B) TELEFAX: (212) 596-9090

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: genomic library in pBluescriptII SK

(B) CLONE: pNB2

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 957..2867

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAGTTTTGCG TTCCCGAACC GAACGTGATG CTTGAGCCGA ACACCTGTCC GGCAAGGCGG	60
CTGACCGCCC CCTTTTGCCC CATCGACATC GTAACAATCG GTTTGGTGGC AAGCTCTTTC	120
GCTTTGAGCG TGGCAGAAAG CAAAGTCAGC ACGTCTTCCG CGCTTTGCGG CATCACCGCA	180
ATTTTGCAGA TGTCCGCGCC GCAGTCCTCC ATCTGTTTCA GACGGCATAAC GATTTCTTCT	240
TGCGGCGGCG TCGGGTGAAA CTCATGATTG CAGAGCAGGG CGGCGATGCC GTTTTTTTGA	300
GCATGCGCCA CGGCGCGCCG GACGGCGGTT TCGCCGAAA AAAGCTCGAT ATCGATAATG	360
TCGGGCAGGC GGCTTTCAAT CAGCGAGTCG AGCAGTTCAA AATAATAATC GTCCGAACAC	420
GGGAACGAGC CGCCTTCGCC ATGCCGTCTG AACGTAAACA GCAGCGGCTT GTCGGGCAGC	480
GCGTCGCGGA CGGTCTGCGT GTGGCGCAAT ACTTCGCCGA TGCTGCCCCG GCATTCCAAA	540
AAATCGGCGC GGAACCTCGAC GATATCGAAG GGCAGGTTTT TGATTTGGTC AAGTACGGCG	600
GAAAGTACGG CGGCATCGCG GGCGACAAGC GGCACGGCGA TTTTGGTGCG TCCGCTTCCG	660
ATAACGGTGT TTTTGACGGT CAGGCTGGTG TGCATGGCGG TTGTTGCGGC TGAAAGGAAC	720
GGTAAAGACG CAATTATAGC AAAGGCACAG GCAATGTTTC AGACGGCATT TCTGTGCGGC	780
CGGCTTGATA TGAATCAAGC AGCATCCGCA TATCGGAATG CAGACTTGGC ACAAGCCCTG	840
TCTTTTCTAG TCAGTCCGCA GTTCTTGCA TATGATTGCA CGACACGCCC TACACGGCAT	900
TTGCAGGATA CGGCGGCAGA CCGCCGGTCG GAAACTTCAG AATCGGAGCA GGCATC	956
ATG TTG ACC CCC ACG CAG CAA GTC GGT TTG ATT TTA CAG TAC CTC AAA	1004
Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys	
1 5 10 15	
ACA CGC ATC TTG GAC ATC TAC ACG CCC GAA CAG CGC GCC GGC ATC GAA	1052
Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu	
20 25 30	
AAA TCC GAA GAC TGG CGG CAG TTT TCG CGC CGC ATG GAT ACG CAT TTC	1100

Lys	Ser	Glu 35	Asp	Trp	Arg	Gln	Phe 40	Ser	Arg	Arg	Met	Asp 45	Thr	His	Phe	
CCC Pro	AAA Lys 50	CTG Leu	ATG Met	AAC Asn	GAA Glu	CTC Leu 55	GAC Asp	AGC Ser	GTG Val	TAC Tyr	GGC Gly 60	AAC Asn	AAC Asn	GAA Glu	GCC Ala	1148
CTG Leu 65	CTG Leu	CCT Pro	ATG Met	CTG Leu	GAA Glu 70	ATG Met	CTG Leu	CTG Leu	GCG Ala	CAG Gln 75	GCA Ala	TGG Trp	CAA Gln	AGC Ser	TAT Tyr 80	1196
TCC Ser	CAA Gln	CGC Arg	AAC Asn	TCA Ser 85	TCC Ser	TTA Leu	AAA Lys	GAT Asp	ATC Ile 90	GAT Asp	ATC Ile	GCG Ala	CGC Arg	GAA Glu 95	AAC Asn	1244
AAC Asn	CCC Pro	GAT Asp	TGG Trp 100	ATT Ile	TTG Leu	TCC Ser	AAC Asn	AAA Lys 105	CAA Gln	GTC Val	GGC Gly	GGC Gly	GTG Val 110	TGC Cys	TAC Tyr	1292
GTT Val	GAT Asp	TTG Leu 115	TTT Phe	GCC Ala	GGC Gly	GAT Asp 120	TTG Leu	AAG Lys	GGC Gly	TTG Leu	AAA Lys	GAT Asp 125	AAA Lys	ATT Ile	CCT Pro	1340
TAT Tyr	TTT Phe 130	CAA Gln	GAG Glu	CTT Leu	GGT Gly	TTG Leu 135	ACT Thr	TAT Tyr	CTG Leu	CAC His	CTG Leu 140	ATG Met	CCG Pro	CTG Leu	TTT Phe	1388
AAA Lys 145	TGC Cys	CCT Pro	GAA Glu	GGC Gly	AAA Lys 150	AGC Ser	GAC Asp	GGC Gly	GGC Gly	TAT Tyr 155	GCG Ala	GTC Val	AGC Ser	AGC Ser	TAC Tyr 160	1436
CGC Arg	GAT Asp	GTC Val	AAT Asn	CCG Pro 165	GCA Ala	CTG Leu	GGC Gly	ACA Thr	ATA Ile 170	GGC Gly	GAC Asp	TTG Leu	CGC Arg	GAA Glu 175	GTC Val	1484
ATT Ile	GCT Ala	GCG Ala	CTG Leu 180	CAC His	GAA Glu	GCC Ala	GGC Gly	ATT Ile 185	TCC Ser	GCC Ala	GTC Val	GTC Val	GAT Asp 190	TTT Phe	ATC Ile	1532
TTC Phe	AAC Asn	CAC His 195	ACC Thr	TCC Ser	AAC Asn	GAA Glu	CAC His 200	GAA Glu	TGG Trp	GCG Ala	CAA Gln	CGC Arg 205	TGC Cys	GCC Ala	GCC Ala	1580
GGC Gly	GAC Asp 210	CCG Pro	CTT Leu	TTC Phe	GAC Asp	AAT Asn 215	TTC Phe	TAC Tyr	TAT Tyr	ATT Ile	TTC Phe 220	CCC Pro	GAC Asp	CGC Arg	CGG Arg	1628
ATG Met 225	CCC Pro	GAC Asp	CAA Gln	TAC Tyr	GAC Asp 230	CGC Arg	ACC Thr	CTG Leu	CGC Arg	GAA Glu 235	ATC Ile	TTC Phe	CCC Pro	GAC Asp	CAG Gln 240	1676
CAC His	CCG Pro	GGC Gly	GGC Gly	TTC Phe 245	TCG Ser	CAA Gln	CTG Leu	GAA Glu 250	GAC Asp	GGA Gly	CGC Arg	TGG Trp	GTG Val	TGG Trp 255	ACG Thr	1724
ACC Thr	TTC Phe	AAT Asn	TCC Ser 260	TTC Phe	CAA Gln	TGG Trp	GAC Asp 265	TTG Leu	AAT Asn	TAC Tyr	AGC Ser	AAC Asn	CCG Pro 270	TGG Trp	GTA Val	1772

TTC Phe	CGC Arg	GCA Ala 275	ATG Met	GCG Ala	GGC Gly	GAA Glu 280	ATG Met	CTG Leu	TTC Phe	CTT Leu	GCC Ala	AAC Asn 285	TTG Leu	GGC Gly	GTT Val	1820
GAC Asp	ATC Ile 290	CTG Leu	CGT Arg	ATG Met	GAT Asp	GCG Ala 295	GTT Val	GCC Ala	TTT Phe	ATT Ile	TGG Trp 300	AAA Lys	CAA Gln	ATG Met	GGG Gly	1868
ACA Thr 305	AGC Ser	TGC Cys	GAA Glu	AAC Asn	CTG Leu 310	CCG Pro	CAG Gln	GCG Ala	CAC His	GCC Ala 315	CTC Leu	ATC Ile	CGC Arg	GCG Ala	TTC Phe 320	1916
AAT Asn	GCC Ala	GTT Val	ATG Met	CGT Arg 325	ATT Ile	GCC Ala	GCG Ala	CCC Pro	GCC Ala 330	GTG Val	TTC Phe	TTC Phe	AAA Lys	TCC Ser 335	GAA Glu	1964
GCC Ala	ATC Ile	GTC Val	CAC His 340	CCC Pro	GAC Asp	CAA Gln	GTC Val 345	GTC Val	CAA Gln	TAC Tyr	ATC Ile	GGG Gly	CAG Gln 350	GAC Asp	GAA Glu	2012
TGC Cys	CAA Gln 355	ATC Ile	GGT Gly	TAC Tyr	AAC Asn	CCC Pro	CTG Leu 360	CAA Gln	ATG Met	GCA Ala	TTG Leu 365	TTG Leu	TGG Trp	AAC Asn	ACC Thr	2060
CTT Leu 370	GCC Ala	ACG Thr	CGC Arg	GAA Glu	GTC Val	AAC Asn 375	CTG Leu	CTC Leu	CAT His	CAG Gln 380	GCG Ala	CTG Leu	ACC Thr	TAC Tyr	CGC Arg	2108
CAC His 385	AAC Asn	CTG Leu	CCC Pro	GAG Glu	CAT His 390	ACC Thr	GCC Ala	TGG Trp	GTC Val	AAC Asn 395	TAC Tyr	GTC Val	CGC Arg	AGC Ser	CAC His 400	2156
GAC Asp	GAC Asp	ATC Ile	GGC Gly 405	TGG Trp	ACG Thr	TTT Phe	GCC Ala	GAT Asp	GAA Glu 410	GAC Asp	GCG Ala	GCA Ala	TAT Tyr	CTG Leu 415	GGC Gly	2204
ATA Ile	AGC Ser	GGC Gly 420	TAC Tyr	GAC Asp	CAC His	CGC Arg	CAA Gln	TTC Phe 425	CTC Leu	AAC Asn	CGC Arg	TTC Phe 430	TTC Phe	GTC Val	AAC Asn	2252
CGT Arg	TTC Phe 435	GAC Asp	GGC Gly	AGC Ser	TTC Phe	GCT Ala	CGT Arg 440	GGC Gly	GTA Val	CCG Pro	TTC Phe	CAA Gln 445	TAC Tyr	AAC Asn	CCA Pro	2300
AGC Ser	ACA Thr 450	GGC Gly	GAC Asp	TGC Cys	CGT Arg	GTC Val 455	AGT Ser	GGT Gly	ACA Thr	GCC Ala	GCG Ala 460	GCA Ala	TTG Leu	GTC Val	GGC Gly	2348
TTG Leu 465	GCG Ala	CAA Gln	GAC Asp	GAT Asp	CCC Pro 470	CAC His	GCC Ala	GTT Val	GAC Asp	CGC Arg 475	ATC Ile	AAA Lys	CTC Leu	TTG Leu	TAC Tyr 480	2396
AGC Ser	ATT Ile	GCT Ala	TTG Leu	AGT Ser 485	ACC Thr	GGC Gly	GGT Gly	CTG Leu	CCG Pro 490	CTG Leu	ATT Ile	TAC Tyr	CTA Leu	GGC Gly 495	GAC Asp	2444
GAA Glu	GTG Val	GGT Gly 500	ACG Thr	CTC Leu	AAT Asn	GAC Asp	GAC Asp 505	GAC Asp	TGG Trp	TCG Ser	CAA Gln	GAC Asp 510	AGC Ser	AAT Asn	AAG Lys	2492

AGC GAC GAC AGC CGT TGG GCG CAC CGT CCG CGC TAC AAC GAA GCC CTG 2540
 Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu
 515 520 525

TAC GCG CAA CGC AAC GAT CCG TCG ACC GCA GCC GGG CAA ATC TAT CAG 2588
 Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln
 530 535 540

GGC TTG CGC CAT ATG ATT GCC GTC CGC CAA AGC AAT CCG CGC TTC GAC 2636
 Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp
 545 550 555 560

GGC GGC AGG CTG GTT ACA TTC AAC ACC AAC AAC AAG CAC ATC ATC GGC 2684
 Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly
 565 570 575

TAC ATC CGC AAC AAT GCG CTT TTG GCA TTC GGT AAC TTC AGC GAA TAT 2732
 Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr
 580 585 590

CCG CAA ACC GTT ACC GCG CAT ACC CTG CAA GCC ATG CCC TTC AAG GCG 2780
 Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala
 595 600 605

CAC GAC CTC ATC GGT GGC AAA ACT GTC AGC CTG AAT CAG GAT TTG ACG 2828
 His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr
 610 615 620

CTT CAG CCC TAT CAG GTC ATG TGG CTC GAA ATC GCC TGA CGCACGCTTC 2877
 Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala *
 625 630 635

CCAAATGCCG TCTGAACCGT TTCAGACGGC ATTTGCG 2914

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 637 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys
 1 5 10 15

Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu
 20 25 30

Lys Ser Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe
 35 40 45

Pro Lys Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala
 50 55 60

Leu Leu Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr

[illegible]

His Asn Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His
 385 390 395 400
 Asp Asp Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly
 405 410 415
 Ile Ser Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn
 420 425 430
 Arg Phe Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro
 435 440 445
 Ser Thr Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Ala Leu Val Gly
 450 455 460
 Leu Ala Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr
 465 470 475 480
 Ser Ile Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp
 485 490 495
 Glu Val Gly Thr Leu Asn Asp Asp Asp Trp Ser Gln Asp Ser Asn Lys
 500 505 510
 Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu
 515 520 525
 Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln
 530 535 540
 Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp
 545 550 555 560
 Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly
 565 570 575
 Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr
 580 585 590
 Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala
 595 600 605
 His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr
 610 615 620
 Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala *
 625 630 635

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCACCATGG GCATCTTGGA CATC

24

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Neisseria polysaccharea*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CTGCCATGGT TCAGACGGCA TTTGG

25

Neisseria polysaccharea